# Agreement in Self–Other Ratings of Leader Effectiveness: The role of demographics and personality

**Robert P. Vecchio* and Robert J. Anderson****

*Department of Management, Mendoza College of Business, University of Notre Dame, Notre Dame, 46556 IN, USA. Vecchio.1@nd.edu
**The Leadership Circle, 10918 Springbrook Court, Whitehouse, OH 43571, USA. randerson@theleadershipcircle.com

**Personality and demographic attributes for a set of 1221 focal managers were examined as correlates of leadership effectiveness evaluations that were obtained via a 360° feedback program. Polynomial regression was used to study the congruence of self-ratings provided by focal managers relative to the different evaluative perspectives (i.e., immediate superior, peer, and subordinate). Analyses supported the prediction that focal manager's sex and age would be associated with the ratings provided by themselves and others. Plus, the tendency to overestimate one's own leader effectiveness relative to evaluations provided by others was found to be greater for males and older managers. Focal managers who expressed greater social sensitivity were evaluated more favorably by subordinates and peers, although not by superiors. Ratings of leader effectiveness from immediate superiors were, instead, more readily predicted by judgments of the performance of the focal manager's organizational unit relative to comparable units. Results of polynomial regression analysis, however, indicated that self–other agreement was related to the focal's sex, social sensitivity, and social dominance. Implications for understanding obstacles to openness to change are discussed.**

## 1. Introduction

Because actual numerical disagreement (discrepancy) between self and other appraisals provides a basis for discussion in feedback sessions with focal managers, self–other agreement is a key concern for the study of 360° feedback programs. Of particular interest is the tendency of some focal managers to engage in overestimation of their personal effectiveness. Overestimation can be especially problematic as those focal managers who have self-appraisals that exceed the appraisals offered by others are likely to be more resistant to constructive feedback concerning how they are perceived and the need to enact change in their own behavior (Brett & Atwater, 2001). Underestimation is arguably of some importance as well. However, underestimation is thought to be less of a practical concern as focal managers who provide underestimated self-appraisals are likely to express pleasant surprise on seeing the appraisals offered by others and are likely to feel that less criticism will be directed toward them.

The topic of multi-source feedback has also been the subject of increasing research interest as a consequence of the growth of 360° feedback programs (Bailey & Fletcher, 2002; Bracken, Timmreck, & Church, 2001; Church, 2000; Goldsmith, Lyons, & Freas, 2000; Hedge, Borman, & Birkeland, 2001). Of special interest is whether multiple-source appraisals are influenced by such extraneous factors as demographic and personality attributes. Because self-ratings are widely regarded with suspicion due to the likelihood of being inflated,

appraisals obtained from others in the workplace are thought to offer significant added value (Brett & Atwater, 2001; DeNisi & Kluger, 2000; Kluger & DeNisi, 1996). Moreover, there is considerable evidence that points to self-ratings being inflated/biased (Arnold & Davey, 1992; Ashford, 1989; Carless & Roberts-Thompson, 2001; Dunning, Heath, & Suls, 2004; Podsakoff & Organ, 1986; Mabe & West, 1982; Murphy & Cleveland, 1991, 1995; Yammarino & Atwater, 1993, 1997). If there are gaps between self-appraisals and the appraisals obtained from others, then differences in these appraisals may pose a potential problem for a focal manager (Atwater, Brett, & Charles, 2007; Atwater, Waldman, Atwater, & Cartier, 2000; Atwater & Yammarino, 1997a, 1997b).

Before great faith can be placed in multi-source appraisals, it is important to determine whether these appraisals are prone to personal biases. Research on multi-source ratings has found only modest agreement among sources in their appraisals of a focal person (Darr & Catano, 2008; Harris & Schaubroeck, 1988; Smither, London, & Reilly, 2005). Plus, differences in appraiser ratings have been found to be stable across time and across appraisal instruments (Nilsen & Campbell, 1993). Evidence on the structural equivalence of constructs across rater types (Hannum, 2007) also suggests only marginally adequate equivalence. While different sources of appraisal can potentially provide unique information from alternative perspectives (Conway & Huffcutt, 1997), it is still critical to estimate whether these sources are influenced by potential rater bias that may be associated with individual differences among focal managers.

Studies that have explicitly examined the role of individual differences in multi-source appraisals have focused attention on demographic and personality attributes of focal managers (Brutus, Fleenor, & McCauley, 1999; Church & Waclawski, 2001; Fletcher & Baldry, 2000; London & Wohlers, 1991; Ostroff, Atwater, & Feinberg, 2004). These two sources of individual differences attributes are thought to (a) be related to the way that focal managers view themselves (because of differences in their life experiences and values), and (b) elicit tendencies of others to presumptively judge the focal manager.

## 1.1. Demographics: sex, age, race

Research on sex differences in rating appraisals suggests that females tend to rate themselves lower than males (Ostroff *et al*., 2004; Pazy, 1986; Wohlers & London, 1989), and should therefore be inclined to underestimate their own performance (and conversely, males should reveal a tendency to overestimate their own performance). Older focal managers may also be subject to providing inflated self-appraisals (Ostroff *et al*., 2004; Wohlers, Hall, & London, 1993). This inflationary tendency may be due to (a) a self-serving bias wherein greater experience is presumed to be an asset, as well as (b) a maturational dynamic wherein self-confidence increases with age for normal adults (Gove, Ortega, & Style, 1989; Lall, Jain, & Johnson, 1996; Robins, Trzesniewski, Gosling, Tracy, & Potter, 2002). Also, unfavorable age-related assumptions about managerial performance and societal tendencies to regard increased age as unattractive, or even 'repulsive' (Rosenbaum, 1986), should result in lower ratings from others (i.e., raters of all ages are likely to reveal a preference for younger focals). These dual, opposite tendencies (of older focal managers to provide more positive self-appraisals, and of alternative evaluators to provide more negative appraisals) should increase the likelihood of detecting greater overestimation by older focal managers (Ostroff *et al*., 2004). Research on race and ratings has tended to show that Whites receive higher ratings than those of other racio-ethnic groupings (Roberson & Block, 2001). These differences may reflect negative stereotyping of non-Whites and/or mixed signals from alternative sources that lead to non-Whites having less clarity concerning performance goals and their attainment. These lower and inconsistent ratings provided by others to non-Whites should result in non-Whites being more likely to overestimate their effectiveness (Ostroff *et al*., 2004).

Therefore, we predict that:

***H1*** *Self-evaluations by the focal manager will be correlated with focal manager's sex, age, and race whereby evaluations will be higher for managers who are male, older, and White; and*

***H2*** *Self–other discrepancy in evaluation will be correlated with sex, age, and race whereby overestimation of effectiveness by focal managers will be greater for managers who are male, older, and non-White*.

## 1.2. Personality: social sensitivity and social dominance

Although there is good evidence that employee personality is related to job performance (Barrick & Mount, 1991) and that openness to feedback is related to personality differences (Atwater & Brett, 2005; Funderberg & Levy, 1997; Maurer & Palmer, 1999; McCarthy & Garavan, 2007; Smither *et al*., 2005), comparatively little research has been devoted to studying individual differences among managers on personality dimensions relative to self–other agreement (Atwater *et al*., 2000). In one study, Brutus *et al*. (1999) found that self and other evaluations were positively

related to focals' reports of empathy and social dominance. In the present study, we draw on the findings of Brutus and colleagues and substantial prior research in the area of leadership (dating back to the Ohio State Leadership Studies) which suggests that managerial consideration should be related to appraisals obtained from subordinates (Judge, Piccolo, & Ilies, 2004; Wagner & Harter, 2006). However, focal manager consideration does not seem likely to have a comparable impact on the appraisal ratings provided by superiors. Instead, it may be that a focal's superior would regard displays of social sensitivity as counter to maintaining a dispassionate demeanor toward others in the work setting. Hence, the association of focal manager social sensitivity with subordinate and peer appraisals should be positive, while the association of social sensitivity with superior appraisals should be substantially diminished (i.e., revealing an interaction).[1]

In contrast to social sensitivity, a superior may place greater value on a focal manager's tendency to be socially dominant (i.e., a tendency to be focused on maintaining control and obtaining desired outcomes from subordinates). Such tendencies, however, are not likely to be well-received by subordinates and peers. Therefore, the association of social dominance with favorable appraisal can be predicted to be positive for superiors, but inverse for subordinates and peers. Furthermore, discrepancies between self and others may be greater for focal managers who are relatively socially dominant, as more highly dominant managers may be inclined to provide upwardly biased self-appraisals that will predispose the identification of overestimation relative to other evaluators, while socially sensitive focal managers may be inclined to provide downwardly biased self-appraisals that will predispose the identification of underestimation relative to other evaluators.[2]

***H3*** *The focal manager characteristic of (a) social dominance will be inversely correlated with effectiveness as assessed by subordinates and peers, while the characteristic of (b) social sensitivity will be positively correlated with effectiveness for subordinates and peers.*

***H4*** *Superiors will provide evaluations that are (a) positively associated with the focal's social dominance and (b) inversely associated with the focal's social sensitivity, relative to subordinates and peers.*

In summary, the present study sought to replicate evidence of associations of demographic attributes with congruence in self–other appraisals, within a 360° feedback context. Also, a major goal of the study was to determine whether managerial personality attributes were differentially associated with appraisals from others, such that the same personality attribute may be judged favorably or unfavorably by specific types of appraisers. Evidence in support of these hypothesized tendencies has implications for the interpreting of multi-source feedback and the training of feedback counselors.

### 1.3. Analytic approach

A substantial amount of prior research on self–other agreement in multi-source appraisal has relied on simple difference scores for calibrating discrepancy (or agreement). From a purely practical standpoint, difference scores are of considerable interest as a consequence of developmental feedback sessions utilizing these differences as the basis for discussion and targeting areas of concern. Within the realm of academic research, the use of difference scores has been severely critiqued (Edwards & Parry, 1993; Edwards, 1994, 1995, 2001, 2002; Philips & Bedeian, 1994; Smith & Tisak, 1993; Tisak & Smith, 1994a, 1994b). At the present time, it seems most appropriate to study self–other agreement with polynomial regression, outlined by Edwards (1995), as (e.g.) polynomial regression avoids confounding the effects of component measures. However, it can be instructive to compare the results of analyses of straightforward difference scores (which would be the basis for discussion in managerial feedback sessions) with the results of polynomial regression analyses in order to identify where the two analytic approaches might suggest different conclusions. With the exception of a few important comprehensive studies in this vein (e.g., Atwater, Ostroff, Yammarino, & Fleenor, 1998; Ostroff *et al*., 2004), it is relatively rare to find research that has examined a full 360° set of appraisals (i.e., including self, supervisor, peer, and subordinate), along with polynomial regression. Because of the rarity of studies that include all levels of appraisal, there is also a concern that published results on self–other agreement may not hold for all appraisal sources (as each appraisal source, arguably, sees a different facet of the rate). Therefore, it is important for research in this vein to include all major sources of appraisal.

## 2. Method

### 2.1. Participants

Data were collected on 1221 focal managers who participated in a 360° feedback program that employed the Leadership Circle Inventory. The Inventory is administered via a secure website, where controlled access through the use of coded identifiers is used to ensure confidentiality. The purpose of the obtained ratings is fully developmental in nature (i.e., to assist in enhancing managerial effectiveness) and is not in-

tended for evaluative judgments that would be related to such administrative decisions as pay or promotion.

Focal managers were predominately male (61.7%), white (84.1%), and holders of a college degree (87.7%). Mean age was 42.56, with a range of 20–74 years. Sex and race were coded as 0, 1 (female, male and non-white, white, respectively), while age was coded in years. Participants were employed in a wide range of organizations (e.g., education, health care, government, insurance, manufacturing, nonprofit, service, etc.), and included individuals from all levels of management. Education was coded on a seven-point scale (from high school to doctoral degree), and job level was coded on a five-point scale (from staff to top management). Along with self-ratings from the focal managers, ratings of the focal managers were also obtained from the focal's immediate superior, a peer, and a subordinate. Because the number of subordinates and peers varied for focal managers and in order to avoid analytical problems that are associated with non-independence (or nesting) of responses within workgroups (Bliese & Hanges, 2004), the responses of one peer and one subordinate were randomly selected for each focal manager. The random selection of one peer and one subordinate also avoids the potential masking or distortion of peer and subordinate source perceptions that can occur with averaging responses across members of these two groups.

## 2.2. Measures

The Leadership Circle Profile provides multi-source feedback on 29 behavioral dimensions as appraised by self, superiors, peers, and subordinates (Anderson, 2006). The Circle Profile contains a total of 144 items with Likert-format response options. For purposes of the present study, only the theoretically relevant personality subscales from the Circle Profile were examined for each focal manager: Social Dominance ($\alpha = .82$), items = I tend to control others, I dictate rather than influence what others do, I am domineering, I pursue results at the expense of people; and Social Sensitivity ($\alpha = .80$), items = I form warm and caring relationships, I am compassionate, I connect deeply with others. Responses for these items were obtained on nine-point Likert scales with five anchor labels: 5 = Always, 4.5, 4 = Often, 3.5, 3 = Sometimes, 2.5, 2 = Seldom, 1.5, and 1 = Never. Responses to each set of items were averaged for each respondent.

Each focal manager also provided a self-assessment as to his/her leader effectiveness by completing the following five-item scale: Leader Effectiveness ($\alpha = .88$), items = I am satisfied with the quality of leadership that I provide, Overall, I provide very effective leadership, I am an example of an ideal leader, My leadership helps this organization to thrive, I am the kind of leader that others should aspire to become. Responses to this scale were obtained on the aforementioned nine-point Likert scale format. Evaluators also completed a comparable version of this scale, wherein the item wording reflected their assessment of the leadership provided by the focal manager ($\alpha$'s: superiors = .95, peers = .96, subordinates = .96). To determine a focal manager's over (under) estimate of leader effectiveness, the rating of each alternate rating source (i.e., immediate superior, peer, and subordinate) was subtracted from the focal's self-rating of effectiveness. Hence, higher values on this algebraic difference dimension reflect overestimation. For polynomial regression analyses, original mean ratings (i.e., the average for the five effectiveness items) were studied directly (i.e., without calculating differences).

To more directly assess the focal's unit performance, the immediate superior of the focal manager responded to the following five-point response scale item: please rate the performance of the part of the overall organization/company (e.g., work unit, division, region, etc.) that this person has formal responsibility for leading/managing: overall performance: 7 = One of the best – Top 10%; 6 = Much better than other similar organizations – Top 25%; 5 = Better than most other similar organizations – Between the Top 25 and 50%; 4 = About the same as other similar organizations – About 50%; 3 = Lower than other similar organizations – Between the Bottom 25 and 50%; 2 = Much lower than other similar organizations – Bottom 25%; 1 = One of the worst – Bottom 10%. The wording of this item was deliberately constructed so as to be relevant across a wide range of diverse work settings.

In order to estimate the factorial validity of the focal's self-reported constructs of effectiveness, social sensitivity, and social dominance, principal axis factoring was employed, with varimax rotation. Results identified three factors (with eigenvalues greater than 1) accounting for 65.54% of the variance. Item factor loadings are displayed in Table 1. The pattern of observed item loadings (i.e., $>.40$ for predicted loadings, but weaker or near zero for off-factor loadings) is supportive of the pattern that would be expected from the proposed three-factor view. To further demonstrate convergent and discriminant validity for the proposed constructs, procedures outlined by Fornell and Larcker (1981) were employed in conjunction with confirmatory factor analysis (using AMOS 16.0). Internal composite reliability values for self-rated effectiveness, social sensitivity, and social dominance were all acceptable (.76, .89, and .86, respectively). In addition, the average variance extracted for these same constructs all exceeded the minimum required value of .50 (i.e., .85, .79, and .73, respectively). Moreover, the pattern of correlations among the constructs relative to the square roots of

Table 1. Results of principal axis factoring analysis: item loadings

| | Factor | | |
|---|---|---|---|
| | I | II | III |
| Domineering | | | |
| I tend to control others | .05 | .81 | −.14 |
| I have to get my own way | −.10 | .77 | −.02 |
| I dictate rather than influence what others do | −.11 | .72 | −.20 |
| I am domineering | −.02 | .82 | −.09 |
| I pursue results at the expense of people | .04 | .59 | −.30 |
| Social sensitivity | | | |
| I form warm and caring relationships | .20 | −.08 | .84 |
| I am compassionate | .16 | −.25 | .74 |
| I connect deeply with others | .17 | −.23 | .82 |
| Leader effectiveness | | | |
| I am satisfied with the quality of leadership that I provide | .77 | −.12 | .10 |
| Overall, I provide very effective leadership | .84 | −.04 | .17 |
| I am an example of an ideal leader | .83 | −.05 | .17 |
| I am the kind of leader that others should aspire to become | .76 | .06 | .08 |
| My leadership helps this organization to thrive | .85 | −.05 | .13 |

Table 2. Internal consistency reliability (ICR) values and average variance extracted

| Scale | ICR | 1 | 2 | 3 |
|---|---|---|---|---|
| 1. Effectiveness | .76 | .92 | | |
| 2. Sensitivity | .89 | .42 | .89 | |
| 3. Dominance | .86 | −.12 | −.47 | .86 |

*Goodness-of-fit indices for confirmatory factor analytic results*

| Model | $\chi^2$ | *df* | NFI | TLI | CFI | RMSEA |
|---|---|---|---|---|---|---|
| Single-factor | 2414.8 | 54 | .56 | .37 | .57 | .20 |
| Three-factor[a] | 311.9 | 51 | .94 | .93 | .95 | .06 |

*Notes:* Square roots of the AVEs (which are listed on the primary diagonal) exceed all correlations among the constructs (i.e., the off-diagonal values). NFI, normed fit index; TLI, Tucker–Lewis index; CFI, comparative fit index; RMSEA, root mean square error of approximation. [a]Three-factor model includes Effectiveness, Social Sensitivity, and Social Dominance.

the average variances extracted was in accord with a pattern that affirms discriminant/convergent validity (see Table 2, upper portion). The fit of a single-factor model (where all items completed by the focal manager were set to load on a single, general factor) was then contrasted with the fit of a three-factor model (where each item was set to load on the relevant proposed factor). The difference in $\chi^2$ for the three-factor model relative to the single factor model indicated a significant improvement for the proposed multi-factor model (change in $\chi^2 = 2102.9$, $df = 3$, exceeding the critical $\chi^2$ value of 11.3, $p < .01$). The fit indices for the two models (see Table 2, lower portion) further indicated the superiority of a three-factor characterization of the focal managers' responses.[3]

## 3. Results

Table 3 presents the means, standard deviations, and intercorrelations for all ratings, overestimated rating values, demographic variables, personality assessments, and superior's rating of the focal's unit performance. The mean values of the ratings of leadership effectiveness do not suggest that focal self-ratings were inflated relative to the other rating sources. As the tabled means indicate by their negative values, there was a consistent tendency for the focals, on average, to underestimate their own leader effectiveness. This finding contrasts markedly with Harris and Schaubroeck's (1988, p. 55) finding that self-ratings, on average, were over a half standard deviation higher than supervisor ratings and approximately one-quarter standard deviation higher than peer ratings. This suggests that the specific context of 360° feedback may substantially deflate self-appraisals.

The correlations among the leadership ratings ranged from .13 to .29. For the focal managers, their ratings correlated most highly with the ratings of their superiors ($r = .20$, $p < .01$) and less well with the ratings of their peers ($r = .13$, $p < .01$) and subordinates ($r = .16$, $p < .01$). The highest observed association (i.e., peers with superiors) was .29 ($p < .01$), while the lowest observed association (i.e., peers with focal managers) was .13 ($p < .01$). Following Fisher's transformations of *r*–*z*-values (Raghunathan, 2003), the set of correlations was tested for equality and found to be significantly different as a set of six correlation values that would be initially assumed (under a null hypothesis) to represent a common population ($\chi^2$ test value = 23,789.06 vs critical $\chi^2$ value = 13.28, $p < .01$). The pattern of present correlations did replicate the Harris and Schaubroeck (1988, p. 51) observation that the peer–superior association was higher than either the self–superior or the self–peer association.

In order to determine whether the various ratings of leader effectiveness were related to a performance-based estimate of the focal's job performance, the immediate superior's report of the focal's comparative unit performance was correlated with the relevant variables (see bottom row of Table 3). It is important to note that all sources of appraisal of leader effectiveness (peer, self, and supervisor) were positively correlated with this measure. This suggests that all sources of ratings were valid. Although all sources of evaluation were correlated with this index, the highest correlation

was (understandably) with the immediate superior's ratings of leadership effectiveness. Interestingly, the more highly the focal overestimated his/her leader effectiveness relative to the immediate superior, the more negative was the superior's rating of the focal's job performance ($r = -.29$, $p < .01$), thereby suggesting the validity of the overestimation index. Lastly, the superior's performance rating was significantly, and inversely, correlated with the focal's age ($r = -.11$, $p < .05$), indicating that older focals received lower performance ratings.

### 3.1. Hypothesis tests

Hypothesis 1 predicted that focal self-evaluations would be correlated with the focal's sex, age, and race. Notably, the relevant correlation in Table 3 reveals a significant positive association for self-evaluations with age ($r = .09$, $p < .01$). However, it is likely that other associations that are inherent among these individual differences variables may mask predicted associations. Therefore, step-down regression analyses were conducted where each demographic variable was entered into a second regression equation after entering all other demographic variables in a prior regression equation. The results of this step-down procedure reaffirmed the bivariate results (standardized $\beta$'s: sex $= -.01$, NS; age $= .14$, $p < .01$; race $= -.07$, NS). Of note, the significant bivariate association of self-evaluation with education ($r = .10$, $p < .01$) was not reaffirmed with the step-down result ($\beta = -.04$, NS), suggesting that controlling for age eliminated the apparent positive association of education with self-evaluation.

Hypothesis 2 predicted that self–other discrepancy in evaluations would be correlated with the focal's sex, age, and race. While the bivariate associations listed in Table 3 are suggestive of age and sex being correlated with overestimations of effectiveness, these algebraic difference criteria can be challenged as being the proper basis for determining the contributions of self and other ratings in terms of their potential interaction (cf. Edwards, 1995).

Hypotheses 3a and 3b forecasted that social dominance would be inversely associated with others' evaluations provided by both subordinates and peers, while social sensitivity would be positively associated with these other evaluations. For Hypotheses 4a and 4b, this pattern of associations was forecast to be reversed for superiors. The correlations in Table 3 indicate that social dominance was inversely correlated with peer and subordinate evaluations ($r = -.10$ and $-.11$, respectively, $p < .01$), but not significantly correlated with the evaluations of the superiors (supporting Hypothesis 3a, but not Hypothesis 4a). Interestingly, focal self-evaluations were inversely correlated with social dominance ($r = -.12$, $p < .01$), thereby suggesting that focals with a tendency to be socially dominant may

Table 3. Means, SD, and intercorrelations

| | | Mean | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Self-effective | 3.62 | .54 | – | | | | | | | | | | | | | |
| 2. | Super.-effective | 3.98 | .67 | .20** | – | | | | | | | | | | | | |
| 3. | Peer-effective | 3.97 | .72 | .13** | .29** | – | | | | | | | | | | | |
| 4. | Sub.-effective | 4.11 | .78 | .16** | .20** | .20** | – | | | | | | | | | | |
| 5. | Overest-super. | −.30 | .77 | .53** | −.72** | −.16** | −.08** | – | | | | | | | | | |
| 6. | Overest-peer | −.29 | .84 | .53** | −.14** | −.77** | −.05 | .48** | – | | | | | | | | |
| 7. | Overest-sub. | −.43 | .88 | .47** | −.08** | −.06 | −.80** | .40** | .35** | – | | | | | | | |
| 8. | Sex | .62 | .49 | .00 | −.07** | −.14** | −.07 | .05 | .12** | .05 | – | | | | | | |
| 9. | Age | 42.56 | 8.75 | .09** | −.03 | −.05 | .02 | .07* | .09** | .05 | −.01 | – | | | | | |
| 10. | Education | 4.78 | 1.43 | .10** | .03 | .10* | .08 | .04 | −.02 | −.02 | .06 | .21** | – | | | | |
| 11. | Race | .84 | .37 | −.06 | −.01 | −.08* | −.05 | −.03 | .02 | .01 | .03 | .03 | −.13** | – | | | |
| 12. | Job level | 2.71 | 1.35 | .03 | −.03 | .01 | .00 | .04 | .00 | .01 | .02 | .16** | .53** | .03 | – | | |
| 13. | Dominance | 2.46 | .64 | −.12** | −.03 | −.10** | −.11** | −.05 | .01 | .01 | .16** | −.09** | −.04 | .11** | .00 | – | |
| 14. | Sensitivity | 3.82 | .63 | .36** | .06 | .14** | .14** | .21** | .12** | .11** | −.24** | .02 | .03 | −.08* | −.03 | −.35** | – |
| 15. | Unit performance | 5.81 | 1.14 | .11* | .51** | .18** | .20** | −.29** | −.09 | −.06 | −.08 | −.11* | .05 | −.06 | −.02 | −.02 | .02 |

*Notes:* *$p < .05$; **$p < .01$.

have recognized difficulties that they had in interpersonal relations. Social sensitivity was most strongly correlated with focal self-evaluations of leader effectiveness ($r = .36$, $p < .01$) and, as predicted, was positively correlated with peer and subordinate evaluations ($r = .14$, $p < .01$, and $r = .14$, $p < .01$, respectively). The association of social sensitivity and superior evaluations was not significant. This pattern of findings supports Hypothesis 3b, but not Hypothesis 4b.

While the associations for differences are of some practical importance (in that these results indicate what one is likely to encounter in the practice of managerial counseling), the algebraic difference approach does not provide a clear test of the contribution of the components of self–other agreement. To more rigorously test Hypotheses 2–4, the procedures outlined by Edwards (1995) were employed. As detailed by Edwards (1994, 1995) and Edwards and Parry (1993), the use of algebraic difference scores (such as overestimation values) can introduce methodological confounds that may make the obtained results ambiguous and misleading. The proper alternative procedure calls for the use of polynomial regression to test relationships. The general form of the equation to test for relationships is $Z = b_0 + b_1X_1 + b_2X_2 + b_3X_1^2 + b_4X_2^2 + b_5X_1X_2$. This equation allows for tests between variables, an interaction effect, and curvilinear relationships in the squared terms. In the present instance, self-appraisal is treated as $X_1$ and other-appraisal is treated as $X_2$. The variable of interest for relating to these $X_1$ and $X_2$ variables is the $Z$ variable (i.e., individual demographics and personality dimensions). Table 3 presents the results of these polynomial regression analyses for the various self–other contrasts and the variables of interest. In Step 1 of these analyses, demographic control variables (i.e., all other demographics, plus education level and job level) were included. In Step 2, the main effects of self and other appraisal were added; and in Step 3, the squared and interaction terms were included. Of particular interest are the significance levels of the beta weights for the terms in Step 3.[4]

An examination of the tabled unstandardized regression coefficients (see Table 4) indicates that (a) the significant bivariate associations of appraisals and other variables (Table 3) remain significant even after including the control variables and (b) focal social sensitivity has a substantial association with self and other evaluations and with their interaction in that their incremental contributions to variance accounted for (changes in $R^2$) were significant across all sources of evaluator comparison. Perhaps most critically, significant (self × other) interaction terms were identified for the demographic variable of sex with the appraisals provided by the immediate superior and the peer (unstandardized coefficients = −.498 and −.472, $p < .05$, respectively). It is of interest that the association for immediate superior was not identified in

Table 4. Polynomial regression analyses: associating demographics and personality dimensions to self and other appraisals

| | Sex | | | Age | | | Race | | | Social dominance | | | Social sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Superior | Peer | Subord. | Superior | Peer | Subord. | Superior | Peer | Subord. | Superior | Peer | Subord. | Superior | Peer | Subord. |
| Step 1. | | | | | | | | | | | | | | | |
| Sex | – | – | – | −.228 | −.351 | −.672 | −.179 | −.246 | −.169 | .201** | .156** | .169** | −.362** | −.349** | −.357** |
| Age | −.003 | −.005 | −.010 | – | – | – | .021 | .028* | .027+ | −.003 | −.003 | −.004 | .000 | .000 | −.001 |
| Education | .089 | .104 | .054 | 1.110** | 1.029** | .986** | −.560** | −.558** | −.571** | −.011 | −.015 | −.012 | .026 | .030 | .018 |
| Race | −.165 | −.237 | −.162 | 1.091 | 1.577+ | 1.449 | – | – | – | .187** | .163** | .217** | −.138* | −.102 | −.137* |
| Level | −.021 | −.022 | .016 | .358 | .420 | .317 | .380** | .367** | .423** | .005 | .012 | −.006 | −.032 | −.028 | −.024 |
| $\Delta R^2$ | .005 | .007 | .005 | .050** | .048** | .044** | .072** | .071** | .077** | .041** | .029** | .040** | .086** | .076** | .081** |
| Step 2. | | | | | | | | | | | | | | | |
| Self | −.164 | −.045 | −.216 | 1.127+ | 1.301* | 1.357* | −.198 | −.252 | −.175 | −.115** | −.115** | −.157** | .427** | .437** | .429** |
| Superior | −.113 | | | −.692 | | | .062 | | | .011* | | | .018 | | |
| Peer | | −.465** | | | −1.202** | | | −.161 | | | −.045 | | | .079** | |
| Subordinate | | | −.205+ | | | −.555 | | | −.088 | | | −.020 | | | .055+ |
| $\Delta R^2$ | .005 | .031** | .012+ | .007 | .015** | .009+ | .002 | .007 | .003 | .011** | .014** | .019** | .136** | .150** | .136** |
| Step 3. | | | | | | | | | | | | | | | |
| Self$^2$ | .050 | .039 | −.028 | .441 | .600 | .013 | −.582** | −.493** | −.656** | −.156** | −.161** | −.167** | .165** | .131** | .180** |
| Other$^2$ | −.068 | −.040 | −.308** | .258 | .403 | .278 | .170 | .184 | −.264+ | .498 | .037 | .456 | .051 | .007 | .057+ |
| Self × Other | −.498* | −.472* | .095 | .806 | .538 | .709 | .049 | −.313 | .117 | .126+ | .121* | .027 | −.129+ | −.016 | −.056 |
| $\Delta R^2$ | .010 | .010 | .014+ | .003 | .003 | .002 | .027** | .025** | .039** | .018** | .022** | .018** | .018** | .010* | .022** |
| Total $R^2$ | .020 | .049** | .031 | .059** | .066 | .055 | .101** | .103** | .119** | .069** | .064** | .077** | .240** | .236* | .239** |

*Notes:* $^+p < .10$; $^*p < .05$; $^{**}p < .01$.

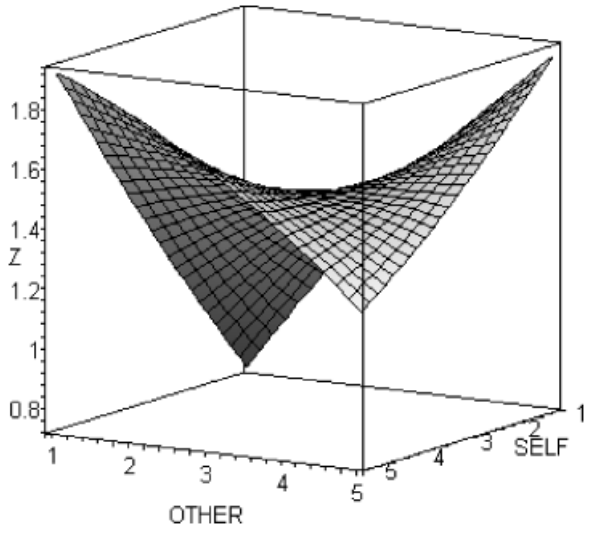


Figure 1. Fitted surface graph examining relations between self and other (immediate superior) appraisal and focal sex.

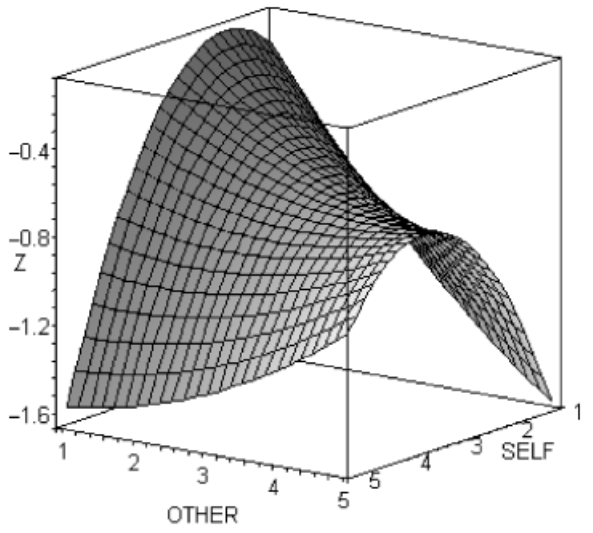


Figure 3. Fitted surface graph examining relations between self and other (immediate superior) appraisal and focal domineering.

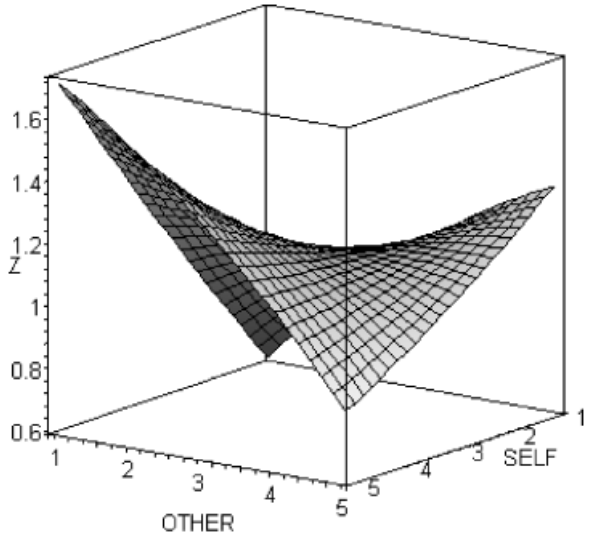


Figure 2. Fitted surface graph examining relations between self and other (peer) appraisal and focal sex.

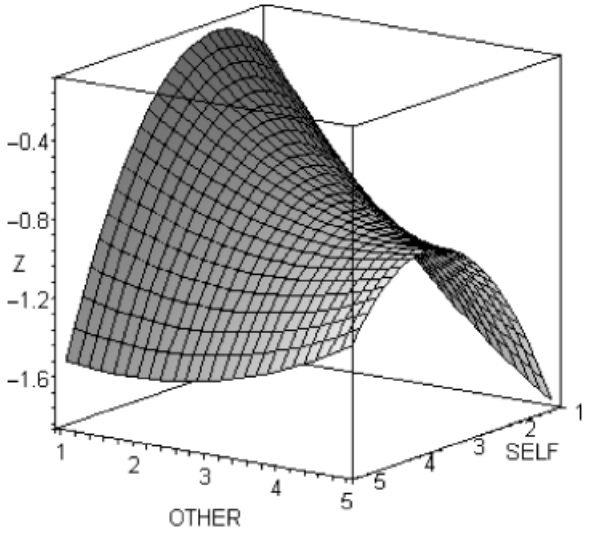


Figure 4. Fitted surface graph examining relations between self and other (peer) appraisal and focal domineering.

the bivariate analysis. For the personality dimension of social dominance, the appraisal provided by the peers was significant in its interaction with self-appraisal (unstandardized coefficient = .121, $p < .05$).

Figure 1 provides an illustration of the fitted surface graph describing the significant interaction relationship for ratings by the focal and the immediate superior relative to focal sex.[5] As indicated in the saddle-shaped response surface graph portrayed in Figure 1, high self-ratings in combination with lower immediate superior ratings were associated with the focal being male (higher on *Z*). Also, higher self-ratings in combination with higher immediate superior ratings were associated with the focal being female. This relationship was inverted when the focals provided lower self-appraisals (i.e., low self and superior appraisals were more associated with being female, and low self and high superior appraisals were associated with being male).

Figure 2 illustrates a similar saddle-shaped response surface for the alternative source of peer-evaluation. In this Figure, the interactive association (which is again readily apparent) reveals that high self-ratings by the focal were in combination with low peer ratings for male focal managers, while high peer ratings for comparable focal managers were associated with being female.

The fitted response surface portrayed in Figure 3 indicates that focals who rated themselves low and who were highly rated by their superior were less socially dominant. Interestingly, a higher level of social dominance was marginally associated with low superior ratings and low self-appraisals. A similar response surface was obtained for the equation for self and peer

appraisals in relation to social dominance (see Figure 4). Again, a lower level of social dominance was related to low self-ratings in combination with high peer ratings, while a higher level of social dominance was associated with low peer ratings and low self-ratings. For both response surfaces portrayed in Figures 3 and 4, an underlying significant negative quadratic coefficient was identified for self-ratings (see Table 4).

Figure 5 illustrates the marginally significant interaction of self and immediate superior ratings for understanding focal manager social sensitivity. The response surface portrayed in Figure 5 reflects an underlying significant positive quadratic coefficient for self-ratings (see Table 4). The graph is interesting as it suggests that focal managers who rate themselves as highly effective and more socially sensitive are rated low by their immediate superior.

Although no formal hypotheses were developed for predicting superior ratings of unit performance, polynomial regression was also used in the interest of completeness and curiosity to estimate the joint effect of superior and self-appraisals for explaining variance in unit performance ratings. Fit (or agreement) between superior and self-appraisals can be mapped in a space defined by these two appraisal dimensions where all points that agree on the surface representing 'superior = self' is referred to as a fit line (where fit can range from low to high). An examination of the shape of the surface along the fit line (in terms of slope and curvature) can provide insights as to the nature of the observed relationship for predicting the criterion. The slope of the fit surface is provided by the sum of the first two regression coefficients given in the earlier stated general form of a polynomial function, while the curvature is given by the sum of the final three regression coefficients. Similarly, the slope of a misfit surface can also be determined by calculating the difference of the first two regression coefficients, while the curvature of the misfit surface can be determined by subtracting the sum of the interaction and final squared terms' coefficients from the initial squared term in the equation.

Table 5 reports the regression coefficients and calculated slopes and curvatures for the fit and misfit

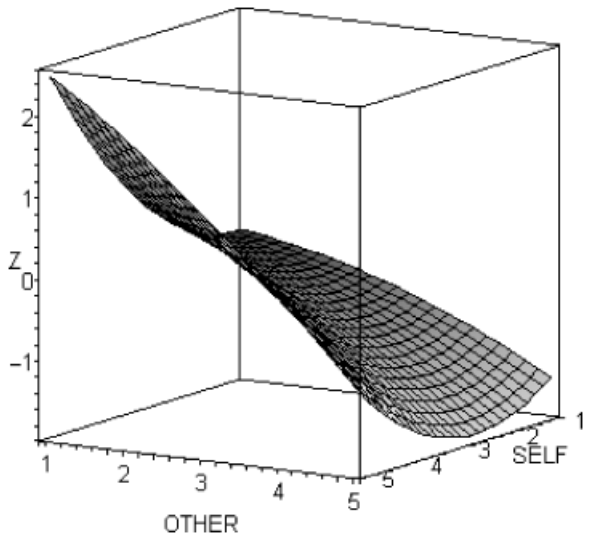


Figure 5. Fitted surface graph examining relations between self and other (immediate superior) appraisal and focal social sensitivity.

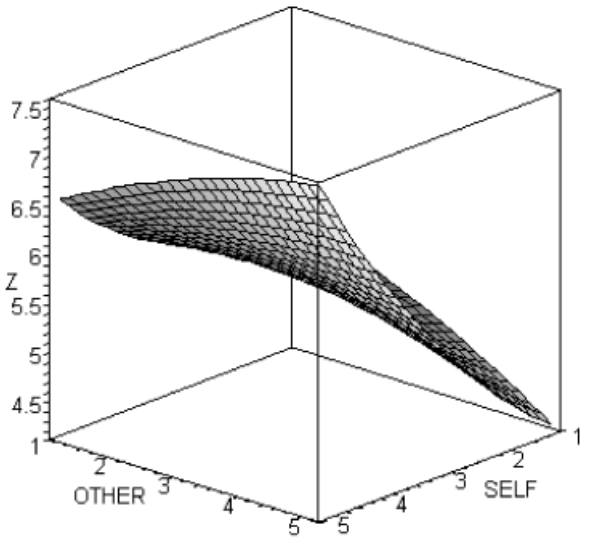


Figure 6. Fitted surface graph examining relations between self and immediate superior appraisal and unit performance rating by superior.

Table 5. Polynomial regression results and examination of response surface features

| Dependent Variable | Unstandardized regression coefficients | | | | | $R^2$ | Fit surface | | Misfit surface | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Superior Rating $b1$ | Self Rating $b2$ | Superior$^2$ $b3$ | Self$^2$ $b4$ | Super. × Self $b5$ | | Slope $(b_1+b_2)$ | Curvature $(b_3+b_5+b_4)$ | Slope $(b_1-b_2)$ | Curvature $(b_3-b_5+b_4)$ |
| Superior's Performance Rating of the Focal's Unit | −.46 | −.28 | .10 | −.03 | .14 | .22** | −.14 | .21 | −.18 | −.07 |

*Note:* $^{**}p<.01$. Tabled results are estimated coefficients after initial step of regression analysis that included the control variables of sex, age, education, race, and job level.

surfaces. Both surfaces revealed a negative slope, while the fit surface demonstrated a positive curvature and the misfit surface demonstrated a slight negative curvature. Although none of the regression coefficients was statistically significant, the associations among the predictor variables and the performance criterion were graphed in the interest of curiosity (see Figure 6). As can be seen in the Figure, high positive appraisal agreement is associated with higher unit performance ratings. However, low performance ratings are tied to lower self-appraisals in conjunction with high superior appraisals. This suggests that self-appraisals may possibly reflect self-perceptions of unit performance to a greater extent than superior appraisals. Or, in other words, superiors are willing to accord focal managers higher ratings for leader effectiveness than unit performance would seem to warrant, while focal self-appraisals may incorporate perceptions of unit performance.[6]

## 4. Discussion

The present findings underscore the value of employing a polynomial regression approach when trying to account for self–other agreement. More specifically, results that were identified by bivariate analyses were not reliably identified with polynomial regression analyses. Hence, the present study makes a methodological contribution in that it is one of the few efforts to study self–other discrepancy with data from a 360 setting in conjunction with the preferred polynomial regression approach. Therefore, the present results provide insights to dynamics that can potentially impede a focal manager's knowledge of discrepancies in self- and other-perceptions, and, thereby, limit openness to change.

The observed tendency for the focal managers to underrate themselves is of interest. As this finding is counter to popular expectations, the result is important in itself. It may be that the developmental context of the ratings predisposed underrating by focals, or it may be that the aggregate findings reported by Harris & Schaubroeck *vis-à-vis* the present finding reflect a shift over time in the tendency of managers to be inclined to underrate themselves. The present results for sex were also somewhat unexpected, and therefore noteworthy, in that males did not rate themselves more highly than females. Females did, however, receive significantly higher evaluations than males from superiors and peers, although not from their subordinates. Males, nonetheless, did reveal a tendency to overestimate their leader effectiveness relative to their superiors' and peers' assessments. This tendency for males to be overraters replicates a finding by Ostroff *et al.* (2004). Also, females in the present study described themselves as significantly more socially sensitive than males ($r = -.24$, $p < .01$), and comparatively less domineering ($r = .16$, $p < .01$), thereby confirming popular, stereotypically based expectations.

Older focal managers did, as hypothesized, provide higher self-evaluations (replicating the findings of Wohlers *et al.*, 1993) and did appear to overestimate their effectiveness relative to their superiors' and peers' assessments (as indicated in the step-down regression results). However, the polynomial regression analyses did not corroborate this suggested tendency toward overestimation. Nonetheless, older focal managers did receive lower ratings from their superiors for their unit performance ($r = -.11$, $p < .01$). This finding corroborates the results reported by Ostroff *et al.* (2004), as well as by van der Heijden (2001).

Race showed little association with the other variables, other than non-Whites describing themselves as significantly more socially sensitive and less domineering than Whites. However, this finding may be partially a function of differences in job demands that may be associated with race rather than with racial differences *per se* (e.g., note that the present correlation of race and education was significant, and that education is commonly related to many job attributes). In the present study, education (e.g.) was strongly related to job level ($r = .53$, $p < .01$). Ostroff *et al.*'s (2004) findings for race, however, indicated that non-White managers were overraters (rating themselves more highly than White managers).

Although overestimation by focal managers was somewhat tied to focal sex, the personality difference of social sensitivity also contributed to predicting the propensity to overestimate one's effectiveness. Plus, focal managers who were socially dominant did not tend to overestimate their leadership consistently in comparison with all other sources. However, self-reports of social sensitivity were strongly associated with self-assessments of leader effectiveness. Assessments of leadership effectiveness provided by both peers and subordinates were associated with the focal's social sensitivity and social dominance, while superior leadership assessments were not. This partially supports the logic that social sensitivity of a focal manager is valued more by peers and subordinates rather than by superiors, while being socially dominant is not judged negatively by superiors. Interestingly, an examination of the response surfaces for understanding social dominance offered a very different image. Instead, the relationship might be more accurately described as highly complex, where low other ratings and self-ratings vary by degree of social dominance. As indicated by the magnitudes of the various correlations, it appears that superiors' leader effectiveness ratings (relative to peer and subordinate ratings) are more readily predicted by estimates of comparative unit performance standings.

While no *a priori* hypothesis was tested, the use of polynomial regression to study superior unit performance ratings revealed interesting results when contrasted with the use of simple algebraic difference

scores to establish overestimation of effectiveness. As indicated by the bivariate correlational results (Table 3), difference score-based overestimates of leader effectiveness (where superior-appraisals were subtracted from self-appraisals) suggested that these overestimates were negatively correlated with unit performance assessments ($r = -.29$, $p < .01$). However, the polynomial regression analyses and an examination of the response surface suggested positive curvature (a *U*-shaped curve) to the surface. This implies that as joint values diverge from the fit line, performance assessments reveal a tendency to decline. This offers an additional perspective for understanding the joint association of appraisal sources when predicting unit performance ratings.

### 4.1. Practical implications

Perhaps the most consistent finding that emerges from the present results is the indication that older focal managers are viewed somewhat negatively. Older managers generally possess greater experience and social standing and hence may have more favorable self-appraisal tendencies. Similarly, they may have higher self-confidence which can lead to an inflated self-appraisal. This suggests that greater tact may be required when providing counseling in feedback sessions for older managers, as their favorable self-image may serve as an obstacle to accepting critical feedback and being open to personal change. Also, this finding suggests that organizations should train appraisers on this topic so as to minimize this potential source of bias.

Focal sex was related to self and other appraisals in a complex manner. The polynomial interactions suggested that males were likely to provide higher self-appraisals in conjunction with low superior and low peer appraisals. Females were likely to receive low superior appraisals in conjunction with low self-appraisals. However, the high points of the saddle in Figures 1 and 2 indicate that discrepancies in appraisals were more commonly found with males. Again, this suggests that resistance to feedback may be greater for male focal managers.

Personality differences were also shown to be related in complex ways to appraisals. Female managers described themselves as significantly more socially sensitive, while male managers described themselves as significantly more socially dominant. Although social dominance and social sensitivity were not associated with immediate superior's rating of unit performance, perceptions of the focal's leader effectiveness did suggest interesting interactions, such that focal managers may be advised to display empathy and a domineering demeanor in a selective manner. Specifically, in the presence of subordinates and peers, social sensitivity is the preferred leadership style, while social dominance should be curtailed. In the presence of one's superior, displaying sensitivity may be detrimental to being perceived as effective (Figure 5), while being socially dominant has no clear relationship. This pattern of results suggests that a Janus-faced style of leadership (wherein a manager modifies his/her style for a specific audience) may be differentially effective across constituencies.

### 4.2. Limitations and future research

To be sure, a single study often contains a variety of unique features that can be highlighted as strengths and limitations. In the present instance, the study's strengths lie in the range of participants (e.g., in terms of a broad sampling of work settings), the unique character of the combination of measures, and the use of polynomial regression to decompose the elements of associations. Additionally, the participants were likely to have been candid in their responses as their goal in providing data was one of self-help through seeking constructive, developmental feedback.

In terms of limitations, the present study included personality measures that are comparatively new (although the present psychometric properties of the measures were shown to be good and prior research on the Leadership Circle Profile indicates that the measured dimensions have good internal reliability and factorial validity, Anderson, 2006). Common method bias may also pose a limit on the present findings (as, e.g., the superiors' ratings took place at the same time as the ratings of leader effectiveness). Also, there are issues of comparability of performance ratings across settings (although the measure was deliberately constructed to maximize comparability via careful construction of the measure's wording). While it is generally believed that different appraisal sources offer different perspectives on a focal's job performance (Conway & Huffcutt, 1997; Mount, Judge, Scullen, Stysma, & Hezlett, 1998), it is of course not possible to rule out completely the possibility that discrepancies in appraisals across sources are due largely to error (Hannum, 2007; Scullen, Mount, & Goff, 2000; Viswesvaran, Schmidt, & Ones, 2002). Therefore, further research into the processes that may moderate the convergence of ratings is warranted.

The purpose of the present ratings may also limit the interpretation of the findings in that the evaluations were obtained for constructive, developmental reasons. Had the ratings involved 'high stake' outcomes (such as pay or promotion), the results could have been quite different. In such settings, underestimation by focal managers may be comparatively uncommon. If a zero-sum distribution of rewards were expected, the ratings provided by others may possibly be influenced in a more negative direction. Future research that compares self–other ratings in distinctly different settings

(i.e., developmental vs evaluative) would be valuable for determining the extent of such contextual influence. To be cautious, the present findings can only be taken as relevant for a developmental context. However, it is important to note that recent research by Zimmerman, Mount, and Goff (2008) showed moderate to strong associations between developmental and administrative multi-source feedback ratings when they were taken from the same rating source on the same performance dimension (p. 129).

A further limitation of the present findings is given by a consideration of the larger societal context in which this evidence was obtained. Specifically, recent cross-cultural findings indicate that the practice of multi-source feedback in the United States may involve different social dynamics than in other parts of the world. For example, Atwater, Waldman, Ostroff, Robie, and Johnson (2005) reported that the simultaneous inclusion of both self and other ratings was generally less useful for five European countries relative to the United States, and Varela and Premeaux (2008) found that multi-source feedback evaluations may be distorted by such cultural values as collectivism and power distance. Other research that compared Western and East Asian countries (Kelly, Whatley, & Worthley, 1990) found that positive leniency in self-appraisal occurred in all of the Western countries, but in none of the East Asian countries.

### *4.3. Conclusions*

Because conflict can arise when self-appraisals differ from feedback provided by others, the study of self–other agreement has important practical implications. Self-ratings are also thought to be vulnerable to several biases, especially leniency and halo. Plus, negative bias may exist in the ratings obtained from others. Judged in their totality, the present results suggest that discrepancy between self and others (over/under estimation) may partially be a result of perceptual propensities tied to specific demographic attributes, and differences in the expectations that various evaluative sources place on the behavior of focal managers. Also, leniency bias was not identified in the present study (which raises the possibility that self-ratings in some 360-feedback programs may activate a self-discounting response from focal managers in order to reduce the likelihood of conflict with others). Further, variance in self–other ratings overlapped with demographic attributes in the range of 1–7%, and with personality variables in the range of 6–24% (suggesting non-trivial associations). Because earlier studies of demographic and personality differences in self–other agreement did not utilize polynomial regression, the present findings also offer new insights as to the interplay among self–other ratings in conjunction with managerial attributes.

## Notes

1. Although there is evidence of sex differences favoring women as excelling in tasks measuring social sensitivity, most of the tasks have measured accuracy in female-relevant domains. When the content domain and the motivational drives are equated to be specifically female-relevant and male-relevant, gender differences are substantially reduced (Hall & Schmid-Mast, 2008; Schmid-Mast & Hall, 2006). Because work settings vary so greatly in terms of relevancy of content and goal motivation with respect to potential gender advantage, no specific gender-related hypotheses are offered (i.e., we may reasonably expect to find that females are higher on social sensitivity or that males are higher on social sensitivity, depending on the overall composition of the sample reflecting a given proportion of gender-favorable settings).
2. Although it is generally acknowledged that males report greater interest in attaining social dominance (Neppl & Murray, 1997; Sidanius, Pratto, & Rabinowitz, 1994), the manipulation of gender identification has been found to moderate this relationship (Wilson & Liu, 2003) and social dominance has not been found to be associated with an individual's sex in any simple fashion (Moskowitz, 1993; Ratliff & Conley, 1981). Hence no predictions are offered for social dominance and focal manager's sex. It is possible, as well, that in some specific contexts (such as in an emergency), social dominance may be especially valued by subordinates and peers (relative to social sensitivity). However, in most work settings and with retrospective appraisals, we expect social sensitivity to be valued to a greater degree than being socially dominant. Similarly, it is possible that social dominance by a focal manager can be a source of interpersonal conflict with the focal's supervisor. Given the prevalent tendency of employees to seek to please their supervisors (rather than fight with them), we expect this to be relatively rare. Nonetheless, the possibility exists that focals who are more socially dominant may receive more negative appraisals from their superiors. Hence, evidence of a negative association for focal social dominance and supervisor appraisal would suggest support for this potential alternative dynamic.
3. As a further demonstration that the personality measures were valid indicators of their respective constructs, responses from 106 MBAs and business upperclassmen were analyzed. Specifically, respondents completed the same personality measures used in the present study along with items that were selected for being recognized as measures of comparable constructs. Scale items were identified in the International Personality Item Pool (Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006). For the construct of social sensitivity, 10 items representing empathy (Jackson Personality Inventory, Jackson, 1994) and eight items representing empathy (Cloninger Temperament and Character Inventory, Cloninger, 1994) were included. For the construct of social

dominance, 10 items representing dominance (California Personality Inventory, Gough & Bradley, 1997) and 10 items representing assertiveness (NEO-PI-R, Costa & McCrae, 1992) were included. The average correlation of the three sensitivity measures was .55, and the average correlation of the three dominance measures was .50. The average correlation of the sensitivity and dominance scales was −.20. The alphas of the six scales ranged from .70 to .87. Confirmatory factor analytic results for this supplemental sample (available from the authors) further corroborated the reasonableness of treating the present measures of social sensitivity and social dominance as valid indicators of the proposed constructs.

4. Conceptually, it seems reasonable to treat multiple individual differences demographics and personality dimensions, and the ratings measures, as variable sets in a canonical correlation analysis. However, Edwards (1995, pp. 320–321) has identified a number of serious limitations associated with such an approach. It is also important to note that in the present regression analyses, the demographic and personality variables are treated as criteria and the components of agreement as predictors. Because the present focus is simply on associations and not causal effects, the notation of variables as predictors or criteria is essentially arbitrary. As focal manager sex and race are dichotomous variables, logistic regression was employed for the relevant analyses.
5. Additional contour graphs are available from the authors on request.
6. While it may seem initially somewhat odd for a superior to rate a focal as highly effective while also rating the focal's unit performance as comparatively low, it is worth noting that a focal manager can be doing an admirable job in an otherwise difficult circumstance (e.g., leading a struggling unit, trying to introduce change to invigorate a unit, coping with a dominant competitor in a given area, etc.). The ability of the superiors to make such a distinction in the ratings indicates that the unit performance ratings were not simply redundant with the effectiveness ratings.

## References


Anderson, R.J. (2006) The Leadership Assessment Circle Profile. *Industrial & Commercial Training*, **38**, 175–184.

Arnold, J. and Davey, K.M. (1992) Self-Ratings and Supervisor Ratings of Graduate Employees' Competences During Early Career. *Journal of Occupational and Organizational Psychology*, **65**, 235–250.

Ashford, S.J. (1989) Self-Assessments in Organizations: A literature review and integrative model. *Research in Organizational Behavior*, **11**, 133–174.

Atwater, L.E. and Brett, J.F. (2005) Antecedents and Consequences of Reactions to Developmental 360 Degree Feedback. *Journal of Vocational Behavior*, **66**, 532–548.

Atwater, L.E., Brett, J.F. and Charles, A.T. (2007) Multisource Feedback: Lessons learned and implications for practice. *Human Resource Management*, **46**, 285–307.

Atwater, L.E., Ostroff, C., Yammarino, F.J. and Fleenor, J.W. (1998) Self–Other Agreement: Does it really matter? *Personnel Psychology*, **51**, 577–598.

Atwater, L.E., Waldman, D., Atwater, D. and Cartier, P. (2000) An Upward Feedback Field Experiment: Supervisors' cynicism, follow-up and commitment to subordinates. *Personnel Psychology*, **53**, 275–297.

Atwater, L.E., Waldman, D., Ostroff, C., Robie, C. and Johnson, K.M. (2005) Self–Other Agreement: Comparing its relationship with performance in the U.S. and Europe. *International Journal of Selection and Assessment*, **13**, 25–40.

Atwater, L.E. and Yammarino, F.J. (1997a) Self–Other Rating Agreement: A review and model. In: Ferris, G. (ed.), *Research in Personnel and Human Resources Management*, Vol. 15. Greenwich, CT: JAI Press, pp. 121–174.

Atwater, L.E. and Yammarino, F.J. (1997b) Does Self–Other Agreement on Leadership Perceptions Moderate the Validity of Leadership and Performance Predictions? *Personnel Psychology*, **45**, 141–164.

Bailey, C. and Fletcher, C. (2002) The Impact of Multiple Source Feedback on Management Development: Findings from a longitudinal study. *Journal of Organizational Behavior*, **22**, 775–788.

Barrick, M. and Mount, M. (1991) The 'Big-Five' Personality Dimensions and Job Performance. *Personnel Psychology*, **44**, 1–26.

Bliese, P.D. and Hanges, P.J. (2004) Being too Liberal and too Conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods*, **7**, 400–417.

Bracken, D.W., Timmreck, C.W. and Church, A.H. (2001) *The Handbook of Multisource Feedback: The comprehensive resource for designing and implementing MSF processes*. San Francisco, CA: Jossey-Bass.

Brett, J.F. and Atwater, L. (2001) 360-Degree Feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, **86**, 930–942.

Brutus, S., Fleenor, J.W. and McCauley, C.D. (1999) Demographic and Personality Predictors of Congruence in Multi-Source Ratings. *Journal of Management Development*, **18**, 417–435.

Carless, S.A. and Roberts-Thompson, G.F. (2001) Self-Ratings in Training Programs: An examination of level of performance and the effects of feedback. *International Journal of Selection and Assessment*, **9**, 217–225.

Church, A.H. (2000) Do higher Performing Managers Actually Receive Better Ratings? A Validation of Multirater Assessment Methodology. *Consulting Psychology Journal: Practice and Research*, **52**, 99–116.

Church, A.H. and Waclawski, J. (2001) Hold the line: An examination of line vs. staff differences. *Human Resource Management*, **40**, 21–34.

Cloninger, C.R. (1994) *The Temperament and Character Inventory (TCI): A guide to its development and use*. Washington University, St. Louis, MO: Center for the Psychology of Personality.

Conway, J.M. and Huffcutt, A.I. (1997) Psychometric Properties of Multisource Performance Ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, **10**, 331–360.

Costa, P.T. and McCrae, R.R. (1992) *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Darr, W. and Catano, V.M. (2008) Multisource Assessments of Behavioral Competence and Selection Interview Performance. *International Journal of Selection and Assessment*, **16**, 68–72.

DeNisi, A. and Kluger, A. (2000) Feedback Effectiveness: Can 360-degree appraisals be improved? *Academy of Management Executive*, **14**, 129–139.

Dunning, D., Heath, C. and Suls, J.M. (2004) Flawed Self-Assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, **5**, 69–106.

Edwards, J.R. (1994) Regression Analysis as an Alternative to Difference Scores. *Journal of Management*, **20**, 683–689.

Edwards, J.R. (1995) Alternatives to Difference Scores. *Organizational Behavior and Human Decision Processes*, **64**, 307–324.

Edwards, J.R. (2001) Ten Difference Score Myths. *Organizational Research Methods*, **4**, 265–287.

Edwards, J.R. (2002) Alternatives to Difference Scores: Polynomial regression analysis and response surface methodology. In: Drasgow, F. and Schmitt, N. (eds.), *Measuring and Analyzing Behavior in Organizations: Advances in measurement and data analysis*. Francisco, CA: Jossey-Bass.

Edwards, J.R. and Parry, M.E. (1993) On the Use of Polynomial Regression Equations as an Alternative to Difference Scores in Organizational Research. *Academy of Management Journal*, **36**, 1577–1613.

Fletcher, C. and Baldry, C. (2000) A Study of Individual Differences and Self-Awareness in the Context of Multisource Feedback. *Journal of Occupational and Organizational Psychology*, **73**, 303–319.

Fornell, C. and Larcker, D.F. (1981) Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, **18**, 39–50.

Funderberg, S. and Levy, P.E. (1997) The Influence of Individual and Contextual Variables on 360-Degree Feedback System Attitudes. *Group and Organizational Studies*, **22**, 210–230.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R. and Gough, H.G. (2006) The International Personality Item Pool and the Future of Public-Domain Personality Measures. *Journal of Research in Personality*, **40**, 84–96.

Goldsmith, M., Lyons, L. and Freas, A. (2000) *Coaching for Leadership*. San Francisco, CA: Jossey-Bass.

Gough, H.G. and Bradley, P. (1997) *CPI Manual* (3rd edn). Palo Alto, CA: Consulting Psychologists Press.

Gove, W.R., Ortega, S.T. and Style, C.B. (1989) The Maturational and Role Perspectives on Aging and Self Through the Adult Years: An empirical evaluation. *American Journal of Sociology*, **94**, 1117–1145.

Hall, J.A. and Schmid-Mast, M. (2008) Are Women Always more Interpersonally Sensitive than Men? Impact of Goals and Content Domain. *Personality and Social Psychology Bulletin*, **34**, 144–155.

Hannum, K. (2007) Measurement Equivalence of 360-Degree Assessment Data: Are different raters rating the same constructs? *International Journal of Selection and Placement*, **9**, 293–301.

Harris, M.M. and Schaubroeck, J. (1988) A Meta-Analysis of Self–Supervisor, Self–Peer, and Peer–Supervisor Ratings. *Personnel Psychology*, **41**, 43–62.

Hedge, J.W., Borman, W.C. and Birkeland, S.A. (2001) History and Development of Multisource Feedback as a Methodology. In: Bracken, J.W., Timmereck, C.W. and Church, A.H. (eds.), *The Handbook of Multisource Feedback*. San Francisco, CA: Jossey Bass, pp. 15–32.

Jackson, D.N. (1994) *Jackson Personality Inventory – Revised Manual*. Port Huron, MI: Sigma Assessment Systems.

Judge, T.A., Piccolo, R.F. and Ilies, R. (2004) The Forgotten Ones? The Validity of Consideration and Initiating Structure in Leadership Research. *Journal of Applied Psychology*, **89**, 36–51.

Kelly, L., Whatley, A. and Worthley, R. (1990) Self-Appraisal, Life Goals, and National Culture: An Asian–Western comparison. *Asia Pacific Journal of Management*, **7**, 41–58.

Kluger, A. and DeNisi, A. (1996) The Effects of Feedback Interventions on Performance: A historical review, a meta-analysis, and a preliminary feedback theory. *Psychological Bulletin*, **119**, 254–284.

Lall, R., Jain, V.K. and Johnson, W.B. (1996) Contemporary Norms for the Coopersmith Self-Esteem Inventory– Adult Form. *Perceptual and Motor Skills*, **82**, 1136–1138.

London, M. and Wohlers, A.J. (1991) Agreement Between Subordinate and Self-Ratings in Upward Feedback. *Personnel Psychology*, **44**, 375–390.

Mabe, P. and West, J. (1982) Validity of Self-Evaluation of Ability: A review and meta-analysis. *Journal of Applied Psychology*, **67**, 280–296.

Maurer, T.J. and Palmer, J.K. (1999) Management Development Intentions Following Feedback: Role of perceived outcomes, social pressures, and control. *Journal of Management Development*, **18**, 733–751.

McCarthy, A.M. and Garavan, T.N. (2007) Understanding Acceptance of Multisource Feedback for Management Development. *Personnel Review*, **36**, 903–917.

Moskowitz, D.S. (1993) Dominance and Friendliness: On the interaction of gender and situation. *Journal of Personality*, **61**, 387–409.

Mount, M.K., Judge, T.A., Scullen, S.E., Stysma, M.R. and Hezlett, S.A. (1998) Trait, Rater, and Level Effects in 360-Degree Performance Ratings. *Personnel Psychology*, **51**, 557–576.

Murphy, K.R. and Cleveland, J.N. (1991) *Performance Appraisal*. Needham Heights: Allyn & Bacon.

Murphy, K.R. and Cleveland, J.N. (1995) *Understanding Performance Appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage Publication.

Neppl, T.K. and Murray, A.D. (1997) Social Dominance and Play Patterns Among Preschoolers: Gender comparisons. *Sex Roles: A Journal of Research*, **36**, 381–393.

Nilsen, D. and Campbell, D.P. (1993) Self–Observer Rating Discrepancies: Once an overrater, always an overrater? *Human Resource Management*, **32**, 265–281.

Ostroff, C., Atwater, L.E. and Feinberg, B.J. (2004) Understanding Self–Other Agreement: A look at rater and ratee characteristics, context, and outcomes. *Personnel Psychology*, **57**, 333–376.

Pazy, A. (1986) The Persistence of Pro-Male Bias Despite Identical Information Regarding Causes of Success. *Organi-*

*zational Behavior and Human Decision Processes*, **38**, 366–377.

Philips, A.S. and Bedeian, A.G. (1994) Leader–Follower Exchange Quality: The role of personal and interpersonal attributes. *Academy of Management Journal*, **37**, 990–1001.

Podsakoff, P.M. and Organ, D.W. (1986) Self-Reports in Organizational Research: Problems and prospects. *Journal of Management*, **12**, 531–544.

Raghunathan, T. (2003) An Approximate Test of Homogeneity of Correlated Correlation Coefficients. *Quality and Quantity*, **37**, 99–110.

Ratliff, E.S. and Conley, J. (1981) The Structure of Masculinity–Femininity: Multidimensionality and gender differences. *Social Behavior and Personality: An International Journal*, **9**, 41–47.

Roberson, L. and Block, C.J. (2001) Racioethnicity and Job Performance: A review and critique of theoretical perspectives on the causes of group differences. *Research in Organizational Behavior*, **23**, 247–325.

Robins, R.W., Trzesniewski, K.H., Gosling, S.D., Tracy, J.L. and Potter, J. (2002) Global Self Esteem Across the Life Span. *Psychology and Aging*, **17**, 423–434.

Rosenbaum, M. (1986) The Repulsion Hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, **51**, 1156–1166.

Schmid-Mast, M. and Hall, J.A. (2006) Women's Advantage at Remembering Others' Appearance: A systematic look at the why and when of a gender difference. *Personality and Social Psychology Bulletin*, **32**, 353–364.

Scullen, S.E., Mount, M.K. and Goff, M. (2000) Understanding the Latent Structure of Job Performance Ratings. *Journal of Applied Psychology*, **85**, 956–970.

Sidanius, J., Pratto, F. and Rabinowitz, J.L. (1994) Gender, Ethnic Status, and Ideological Symmetry: A social dominance interpretation. *Journal of Cross-Cultural Psychology*, **25**, 194–216.

Smith, C.S. and Tisak, J. (1993) Discrepancy Measures of Role Stress Revisited: New perspectives on old issues. *Organizational Behavior and Human Decision Processes*, **56**, 285–307.

Smither, J.W., London, M. and Reilly, R.R. (2005) Does Performance Appraisal Improve Following Multisource Feedback? A Theoretical Model, Meta-Analysis, and Review of Empirical Findings. *Personnel Psychology*, **58**, 33–66.

Tisak, J. and Smith, C.S. (1994a) Rejoinder to Edwards' Comments. *Journal of Management*, **20**, 691–694.

Tisak, J. and Smith, C.S. (1994b) Defending and Extending Difference Score Methods. *Journal of Management*, **20**, 675–682.

van der Heijden, B. (2001) Age and Assessments of Professional Expertise: The relationship between higher level employees' age and self-assessments or supervisor ratings of professional expertise. *International Journal of Selection and Assessment*, **9**, 309–324.

Varela, O.E. and Premeaux, S.F. (2008) Do Cross-Cultural Values Affect Multisource Feedback Dynamics? The Case of High Power Distance and Collectivism in Two Latin-American Countries. *International Journal of Selection and Assessment*, **16**, 134–142.

Viswesvaran, C., Schmidt, F.L. and Ones, D.S. (2002) The Moderating Influence of Job Performance Dimensions on Convergence of Supervisory and Peer Ratings of Job Performance: Unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology*, **87**, 345–354.

Wagner, R. and Harter, J. (2006) *12: The Elements of Great Managing*. Princeton, NJ: Gallup Press.

Wilson, M.S. and Liu, J.H. (2003) Social Dominance Orientation and Gender: The moderating role of gender identity. *British Journal of Social Psychology*, **42**, 187–198.

Wohlers, A.J., Hall, M.J. and London, M. (1993) Subordinates Rating Managers: Organization and biographic correlates of self/subordinate agreement. *Journal of Occupational and Organizational Psychology*, **66**, 263–275.

Wohlers, A.J. and London, M. (1989) Ratings of Managerial Characteristics: Evaluation difficulty, co-worker agreement, and self-awareness. *Personnel Psychology*, **42**, 235–261.

Yammarino, F.J. and Atwater, L.E. (1993) Understanding Self-Perception Accuracy: Implications for human resource management. *Human Resource Management*, **32**, 231–247.

Yammarino, F.J. and Atwater, L.E. (1997) Do Managers see Themselves as Others see them? Implications of Self–Other Rating Agreement for Human Resources Management. *Organizational Dynamics*, **25**, 35–44.

Zimmerman, R.D., Mount, M.K. and Goff, M. (2008) Feedback and Leaders' Goal Performance: Moderating effects of rating purpose, rater perspective, and performance dimension. *International Journal of Selection and Assessment*, **16**, 122–133.